・平均の区間推定（二標本の差の区間推定）

いままでは、一つの母集団に対して標本平均や標本分散の推定を行ってきたが、男女の平均の差などのように、二つの母集団の平均の差を調べたいこともある。ここでは、二つの母集団がそれぞれ母分散が等しい正規分布 $N(\mu_1, \sigma^2), N(\mu_2, \sigma^2)$ に従うとき、母平均の差 $\mu_1 - \mu_2$ の区間推定を行う。

まず、2つの母集団からそれぞれ

$$(x_1, x_2, \cdots, x_n) \ \ , \ \ (y_1, y_2, \ \cdots \ , y_m)$$

と1つ目の母集団から n 個、2つ目の母集団から m 個の標本が無作為に得られたとする。このとき母平均の差 $\mu_1 - \mu_2$ の点推定は

$$\bar{x} - \bar{y} = \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{m}\sum_{j=1}^{m} y_j$$

のように、それぞれの標本平均のを計算した差で求めることができる。次にこの点推定の値を使って区間推定を求める。正規分布の和や差が正規分布に従うことから、$\bar{x} - \bar{y}$ も正規分布に従うことになる。実際に $\bar{x} - \bar{y}$ は母平均 $\mu_1 - \mu_2$、母分散 $\frac{\sigma^2}{n} + \frac{\sigma^2}{m}$ の正規分布 $N(\mu_1 - \mu_2, \frac{\sigma^2}{n} + \frac{\sigma^2}{m})$ に従うことが理論的にわかっている。よって、母分散が既知の場合と未知の場合で次のように区間推定を行うことができる。

・母分散が既知のとき

$\bar{x} - \bar{y}$ が正規分布 $N(\mu_1 - \mu_2, \frac{\sigma^2}{n} + \frac{\sigma^2}{m})$ に従うので、

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}} = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

と変換すれば Z が標準正規分布に従うので、前回同様に次のような z_α を決め、

$$\Pr\{-z_\alpha \leqq Z \leqq z_\alpha\} = 1 - \alpha$$

括弧の中の不等式に実際の値 $\bar{x}, \bar{y}$ を代入し、$\mu_1 - \mu_2$ について解くことで

$$-z_\alpha \leqq \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \leqq z_\alpha$$

$$-z_\alpha \times \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \leqq (\bar{x} - \bar{y}) - (\mu_x - \mu_y) \leqq z_\alpha \times \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$$

$$(\bar{x} - \bar{y}) - z_\alpha \times \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \leqq \mu_x - \mu_y \leqq (\bar{x} - \bar{y}) + z_\alpha \times \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$$

のような区間推定を得る。

・母分散が未知の場合

母分散が異なる場合変換後の分布が複雑なので、母分散が共通 $(\sigma_1^2 = \sigma_2^2 = \sigma^2)$ とする。

母分散 σ^2 の代わりに標本分散 s^2 を使うことを考える。しかし標本分散の計算は母集団ごとに s_x^2, s_y^2 と 2 つ計算できるので、等分散の条件をもとに

$$s^2 = \frac{1}{(n-1)+(m-1)}\left\{(n-1)s_x^2 + (m-1)s_y^2\right\}$$

この値を共通の母分散 σ^2 の推定量とする。

母分散 σ^2 のかわりに s^2 を使うことによって、

$$T = \frac{(\bar{X}-\bar{Y})-(\mu_x-\mu_y)}{S\sqrt{\frac{1}{n}+\frac{1}{m}}}$$

と変換すると t は自由度 $(n-1)+(m-1)$ の t 分布に従うことがわかっている。
後は母集団が 1 つの場合と同様に次のような t_α を求め、

$$\Pr\{-t_\alpha \leqq T \leqq t_\alpha\} = 1-\alpha$$

確率の括弧の中身に $\bar{x}, \bar{y}, s^2$ を代入し $\mu_x - \mu_y$ について解くことで

$$(\bar{x}-\bar{y}) - t_\alpha \times s\sqrt{\frac{1}{n}+\frac{1}{m}} \leqq \mu_x - \mu_y \leqq (\bar{x}-\bar{y}) + t_\alpha \times s\sqrt{\frac{1}{n}+\frac{1}{m}}$$

のような区間推定を得る。

例． 二標本の差の区間推定

ある工場で作られた電球 7 個の寿命を調べたところ、下記のような結果を得た。

76.5, 82.4, 93.0, 78.7, 86.6, 94.2, 82.9 (単位：日)

また、これとは別の工場で作られた電球 6 個の寿命を調べたところ下記のようであった。

77.7, 69.5, 79.6, 68.2, 71.4, 67.4 (単位：日)

このとき、母平均の差 $\mu_x - \mu_y$ の 95%信頼区間を求めよ。

解答例）標本平均と標本分散を計算すると、

標本平均 $\bar{x} = 84.9$、標本分散 $s_x^2 = 45.7$、標本標準偏差 $s_x = 6.76$
標本平均 $\bar{y} = 72.3$、標本分散 $s_y^2 = 26.4$、標本標準偏差 $s_y = 5.14$

母分散が未知なので、共通の母分散の推定量 s^2 を求めると

$$\begin{aligned} s^2 &= \frac{1}{(7-1)+(6-1)}\{(7-1)\times 45.70667 + (6-1)\times 26.384\} \\ &= \frac{1}{11}(274.24 + 131.92) = 36.92364 \qquad (s = 6.076483) \end{aligned}$$

となる。T で変換すると自由度 $(7-1)-(6-1) = 11$ の t 分布に従うことになる。

したがって、$\Pr\{-t_\alpha \leqq T \leqq t_\alpha\} = 0.95 = 1-0.05$ となる t_α は 2.201 となる。よって母平均の 95%信頼区間は

$$-2.201 \leqq \frac{(84.9-72.3)-(\mu_x-\mu_y)}{6.076 \times \sqrt{\frac{1}{7}+\frac{1}{6}}} \leqq 2.201$$

の不等式を解いて、 $5.16 \leqq \mu_x - \mu_y \leqq 20.0$ である。

2019年度 確率統計S　小テスト解答用紙　2019.11.29

・二標本平均の差の区間推定

　消費電力が低くなるように電球に改良を加えた結果、改良前後の消費電力は次の通りであった。

改良前 (x)	:	7.3	7.8	7.2	7.4	8.1
改良後 (y)	:	7.4	6.8	6.3	7.1	

（単位 W）

消費電力の分布は同じ母分散を持つ正規分布であるとするとき、次の問いに答えよ。

1) 標本平均 $(\bar{x}, \bar{y})$・標本分散 (s_x^2, s_y^2) を求めよ。

2) 共通の母分散の推定量 s^2 を求めよ。

$$s^2 = \frac{1}{(n-1)+(m-1)}\{(n-1)s_x^2 + (m-1)s_y^2\}$$

3) 改良前後の平均値の差 $\mu_x - \mu_y$ の９９%信頼区間を求めよ。

2019年度神奈川工科大学 確率統計S 演習問題	学科	学年	組	学　籍　番　号	氏　　　名	

提出先：K3-3309号室前　18番のボックス　　提出期限：１２月　２日（月）１７時頃まで